**BDREPORTER设备静态信息**

1. **数据字典**

**1.1表级**

英文名：dwd\_evt\_bdreporter\_device\_static\_info\_dm

中文名：BDREPORTER设备静态信息

表类型：快照表

ETL调度频率：日表

**1.2字段级**

|  |  |  |  |
| --- | --- | --- | --- |
| 字段名 | 字段类型 | 字段中文名 | 备注 |
| unique\_id | STRING | 唯一编号 |  |
| plmn | STRING | PLMN |  |
| mac\_addr | STRING | mac地址 |  |
| imei | STRING | 设备编号 |  |
| rom\_ver | STRING | ROM版本 |  |
| emui\_ver | STRING | EMUI版本 |  |
| device\_name | STRING | 外部型号 |  |
| opta | STRING | OPTA |  |
| optb | STRING | OPTB |  |
| rec\_time | STRING | 记录时间 |  |
| user\_ip\_addr | STRING | 用户IP地址 |  |
| server\_time | STRING | 用户IP地址 |  |
| etl\_time | VARCHAR(30) | ETL时间 |  |

1. **数据来源**

**2.1．来源于数据打点源表:**

数据打点源表 : ODS\_BDREPORTER\_HI\_ANALYTICS\_HM (采集源表)

**2.2．其他依赖的表:**

dwd\_common\_regex (正则表达式表):规范device\_name,app\_ver

1. **数据逻辑**

**3.1．主体逻辑**

**App打点上报的数据 通过采集服务器收集，收集之后的数据推送到ODS,并通过jar包解析成3张小时表，此表是汇总小时表一天的数据**

**3.2．字段逻辑**

| 字段名 | 处理逻辑 |
| --- | --- |
| unique\_id | 直抽，数据平台添加，拉通作为华为设备唯一标识符 |
| plmn | 直抽，公共陆地移动网络 |
| mac\_addr | 直抽，MAC地址。可作为设备唯一标示符 |
| imei | 设备id,做了格式化。没有做非法imei |
| rom\_ver | 直抽，ROM版本号。判断用户rom版本，根据rom版本计算在网量，升级率等 |
| emui\_ver | 直抽，EMUI版本，判断用户emui版本 |
| device\_name | 直抽，外部型号，用于机型维度数据分析 |
| opta | 直抽，opta 国家码。真实性偏低，覆盖率偏低，命名规则暂不清楚 |
| optb | 直抽，optb 运营商码，。真实性偏低，覆盖率偏低，命名规则暂不清楚 |
| rec\_time | 采集时间 精确到毫秒。数据为时间戳格式，通过dateutil函数经行转换 |
| user\_ip\_addr | 直抽，事件发生ip |
| server\_time | 采集时间 精确到毫秒。用dateutil函数经行转换 |
| etl\_time | ETL时间字段。记录数据入库的服务器时间。 |

**3.3．ETL逻辑**

基本是直抽的逻辑。只有个别字段经行数据标准化的处理。

1. **其他**

**4.1． imei处理过程**

4.1.1标准化：revertDeviceId(deviceIdFormat(imei))

**4.2 性能优化**

**DISTRIBUTE BY T1.unique\_id,T1.pt\_evt; 分布式处理数据，合理分配到reduce任务。减少磁盘的IO**